SCIENCE CHINA
Life Sciences

Tsinghua Special Issue
• RESEARCH PAPER •

January 2014 Vol.57 No.1: 97–106
doi: 10.1007/s11427-013-4589-7

Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS

MEI KunRong[1†], JIN Zhe[2†], REN FangLi[2], WANG YinYing[2], CHANG ZhiJie[2*] & WANG XinQuan[1*]

[1]*Ministry of Education Key Laboratory of Protein Science, Center for Structure Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;*
[2]*State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Medicine, Tsinghua University, Beijing 100084, China*

Received October 14, 2013; accepted November 19, 2013

CREPT and p15RS are two recently identified homologous proteins that regulate cell proliferation in an opposite way and are closely related to human cancer development. Both CREPT and p15RS consist of an N-terminal RPR domain and a C-terminal domain with high sequence homology. The transcription enhancement by CREPT is attributed to its interaction with RNA polymerase II (Pol II). Here we provide biochemical and structural evidence to support and extend this molecular mechanism. Through fluorescence polarization analysis, we show that the RPR domains of CREPT and p15RS (CREPT-RPR and p15RS-RPR) bind to different Pol II C-terminal domain (CTD) phosphoisoforms with similar affinity and specificity. We also determined the crystal structure of p15RS-RPR. Sequence and structural comparisons with RPR domain of Rtt103, a homolog of CREPT and p15RS in yeast, reveal structural basis for the similar binding profile of CREPT-RPR and p15RS-RPR with Pol II CTD. We also determined the crystal structure of the C-terminal domain of CREPT (CREPT-CTD), which is a long rod-like dimer and each monomer adopts a coiled-coil structure. We propose that dimerization through the C-terminal domain enhances the binding strength between CREPT or p15RS with Pol II by increasing binding avidity. Our results collectively reveal the respective roles of N-terminal RPR domain and C-terminal domain of CREPT and p15RS in recognizing RNA Pol II.

CREPT, p15RS, structure, RPR domain, C-terminal domain, RNA Pol II

Citation: Mei KR, Jin Z, Ren FL, Wang YY, Chang ZJ, Wang XQ. Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. Sci China Life Sci, 2014, 57: 97−106, doi: 10.1007/s11427-013-4589-7

Tumor development is highly related to uncontrolled cell growth [1], a process in which oncogenes and tumor suppressor genes are both involved [2]. *CREPT* (cell-cycle related and expression-elevated protein in tumor, another name RPRD1B) and *p15RS* (p15INK4b-related sequence, another name RPRD1A), two recently identified genes, are a pair of highly related genes that have opposite functions in regulating cell proliferation [3,4]. The encoded protein of p15RS is related to p15INK4b, a cyclin-dependent kinase inhibitor, and its expression inhibits cell proliferation [3,5,6]. By contrast, CREPT protein promotes cell proliferation by enhancing transcription of cell cycle related genes including the *CYCLIN D1* and is highly expressed in diverse human tumors [4]. The proposed molecular mechanism of transcription enhancement by CREPT is the formation of a chromatin loop that recycles RNA polymerase II (Pol II) from the termination site to the promoter region of a target gene [4]. Interestingly, Rtt103, an ancestor of both CREPT and p15RS in yeast, was demonstrated to play an important role in the regulation of yeast growth [7]. Recently, Rtt103 was also demonstrated to function in DNA damage response

†Contributed equally to this work
*Corresponding author (email: xinquanwang@mail.tsinghua.edu.cn; zhijiec@tsinghua.edu.cn)

[8].

RNA Pol II has a unique C-terminal domain (CTD) in its largest subunit RPB1, which is composed of multiple tandem repeated heptapeptides with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 with up to 52 repeats in mammals [9–11]. Dynamic and reversible modification of the Pol II CTD is essential for the efficient and accurate completion of transcription cycle [12,13]. The different modification state of CTD, particularly in the positions of Ser2 and Ser5, is closely related to the coordination and recruitment of transcription and mRNA processing factors to RNA Pol II [13–16]. In the early stage of transcription, Ser5 phosphorylation (Ser5P) recruits the mRNA capping enzymes [17–19]. As RNA Pol II progresses into the elongation phase, the level of Ser2 phosphorylation (Ser2P) increases while the level of Ser5P decreases [16]. At the late stage of transcription, Ser2P is the most dominating modification form of CTD in RNA Pol II, leading to increased recruitment of factors responsible for cleavage and polyadenylation [14].

Both CREPT and p15RS consist of an N-terminal RPR domain and a C-terminal domain with high sequence similarity, with only significant sequence variations in the linker region connecting these two domains [4]. The RPR domain is reported in several RNA processing and transcription termination factors and functions as a classical RNA Pol II CTD binding domain [20–23]. Rtt103, a homolog of CREPT and p15RS in yeast, functions as a transcription terminator and also contains a RPR domain [23]. It associates with Pol II in the termination region at the 3′-end of a gene with the RPR domain to trigger the transcription termination of a gene together with Rai1, which favors the torpedo model of transcription termination [7]. Biochemical and structural studies have shown that the RPR domain of Rtt103 directly binds to the Ser2P CTD of RNA Pol II [23]. Our recent study showed that CREPT physically interacts with RNA Pol II [4]. Chromatin immunoprecipitation (ChIP) experiment further indicated that CREPT associates with the 5′ promoter region of the *CYCLIN D1* gene [4], which is different from Rtt103. Therefore, we proposed that CREPT might promote the recycling of RNA Pol II from the terminator to the promoter region, which was supported by the chromosome conformation capture (3C) experiment showing that CREPT promotes the formation of a chromatin loop [4].

Although our previous study has shed light on the transcription enhancement by CREPT, the molecular mechanism underlying the opposite effects of CREPT and p15RS in cell proliferation remains elusive. Structural and biochemical studies of the RPR domain and C-terminal domain in CREPT and p15RS have not been reported yet. In this study, we report the crystal structures of the N-terminal RPR domain of p15RS (p15RS-RPR) and the C-terminal domain of CREPT (CREPT-CTD).

1 Materials and methods

1.1 Protein expression and purification

The coding sequences of human CREPT-RPR domain (residues 1–132) and p15RS-RPR domain (residues 1–131) were cloned into pET22b vector with a C-terminal His tag. *Escherichia coli* BL21 (DE3) cells harboring expression constructs were grown at 37°C in LB medium containing 100 mg L^{-1} ampicillin. Expression of target proteins was induced by addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mmol L^{-1} when the cell A_{600} reached 0.6–0.8. After further incubation at 16°C overnight, the cells were pelleted and re-suspended in a lysis buffer (50 mmol L^{-1} Tris-Cl, 500 mmol L^{-1} NaCl, pH 8.0) and lysed by sonication. After centrifuged at 20000×*g* at 4°C for 1 h, the supernatant was transferred onto a Ni-NTA column, and was washed extensively with washing buffer (50 mmol L^{-1} Tris-Cl, 500 mmol L^{-1} NaCl, 20 mmol L^{-1} imidazole, pH 8.0). The bound His-tagged CREPT-RPR and p15RS-RPR were eluted with elution buffer (50 mmol L^{-1} Tris-Cl, 500 mmol L^{-1} NaCl, 500 mmol L^{-1} imidazole, pH 8.0), and further purified with a Superdex-200 gel filtration column in a running buffer containing 50 mmol L^{-1} Tris-Cl (pH 8.0) and 500 mmol L^{-1} NaCl. CREPT-CTD domain (residues 177–326) and its truncation form (residues 177–280) used for multi-angle light scattering (MALS) analysis were expressed and purified in a similar manner with CREPT-RPR and p15RS-RPR, except the running buffer for Superdex-200 gel filtration column was changed into 50 mmol L^{-1} Tris-Cl (pH 8.0) with 150 mmol L^{-1} NaCl.

CREPT-CTD domain (residues 177–326) prepared for crystallization was cloned into pGEX-6P-1 and expressed as a fusion protein with an N-terminal GST tag. The fusion protein was expressed in *Escherichia coli* BL21 (DE3) induced with 0.5 mmol L^{-1} IPTG at 16°C. The cells were harvested and suspended with lysis buffer containing 50 mmol L^{-1} Tris-Cl (pH 8.0), 500 mmol L^{-1} NaCl and 2 mmol L^{-1} DTT. After sonication, the lysate was centrifuged at 20000×*g* for 1 h, and the supernatant containing GST-CREPT-CTD was transferred to a Glutathione Sepharose 4B column. After extensive wash, the bound fusion protein was digested on column with 3C PreScission protease at 4°C. The released CREPT-CTD without GST tag was collected and further purified by gel filtration chromatography using the Superdex-200 gel filtration column with running buffer containing 50 mmol L^{-1} Tris-Cl (pH 8.0), 500 mmol L^{-1} NaCl and 5 mmol L^{-1} DTT.

1.2 Fluorescence polarization measurement

N-terminally 5,6-carboxyfluorescein (FAM)-labeled unphosphorylated Ser2P, Ser5P and Ser7P CTD peptides were purchased from Scilight Biotechnology, LLC, China. Fluorescence polarization experiments were performed in a

black 384 well assay plate (Corning, USA) and measurements were carried out in a fluorescence spectrometer EnVision (PerkinElmer, USA) in assay buffer containing 50 mmol L^{-1} Tris-Cl (pH 8.0) and 500 mmol L^{-1} NaCl at 25°C. Multiple titrations were performed using fixed concentrations of unphosphorylated Ser2P, Ser5P or Ser7P CTD peptides at 0.1 μmol L^{-1} with increasing concentrations of CREPT-RPR or p15RS-RPR (0.5, 1, 2, 5, 10, 20, 50, 100 and 150 μmol L^{-1}) in a final volume of 20 μL reaction mixture. Triple repeats were set for each individual measurement. Samples were mixed by centrifuging at 1000×*g* for 5 min at 4°C before measuring. In the measurement, samples were excited with vertically polarized light at 477 nm, and both vertical and horizontal emission were recorded at 525 nm. Data was analyzed by Origin8 and fitted to the sigmoidal logistic function.

1.3 Co-immunoprecipitation

CREPT and different truncation plasmids were generated by inserting the PCR-amplified fragments into a pcDNA3.1-Myc vector. CREPT was also inserted into a pcDNA3.1-flag vector to obtain the plasmid used for common immunoprecipitation bait. HEK293T cells were plated in 60 mm dishes the day before transfection. Plasmids (2 μg) were transfected. Cells were lysed 24–36 h after transfection in 900 μL of cell lysis buffer (50 mmol L^{-1} Tris-Cl, 150 mmol L^{-1} NaCl, 50 mmol L^{-1} NaF, 0.5% NP-40, pH 7.5) with protease inhibitors to prepare whole cell lysates. Lysates were mixed with anti-Myc and protein G-agarose beads and incubated at 4°C overnight. Immunoprecipitants and 5% of lysates were analyzed by immunoblotting for the indicated proteins.

1.4 Multiple-angle light scattering (MALS)

The molecular masses of CREPT-CTD and CREPT-CTD_177-280 in solution were determined by MALS using the DAWN HELEOSTM II 18-angle static light-scattering system (Wyatt Technology, USA), connected to an Agilent HPLC, and hooked up with a WTC SEC column (Wyatt Technology, USA). The system was first equilibrated with buffer containing 50 mmol L^{-1} Tris-Cl (pH 8.0) and 150 mmol L^{-1} NaCl for more than 12 h. The equilibrated system was then calibrated with BSA at a concentration of 1 mg mL^{-1}. CREPT-CTD and CREPT-CTD_177-280 were prepared in the aforementioned process and concentrated to 1 mg mL^{-1}. Samples were injected to MALS analyzer at a flow rate of 0.5 mL min^{-1} at 16°C. The molecular mass was calculated using ASTRA5.3.4.14 software (Wyatt Technology, uSA).

1.5 Crystallization and data collection

Proteins were prepared as described above. For p15RS-RPR, protein was concentrated to 23 mg mL^{-1} in 50 mmol L^{-1} Tris-Cl (pH 8.0), 500 mmol L^{-1} NaCl and 5 mmol L^{-1} DTT. Crystals were grown at 18°C by vapor diffusion in sitting drops composed of equal volumes of protein and reservoir solution containing 2.2 mol L^{-1} DL-Malic acid, 0.1 mol L^{-1} Bis-Tris propane, pH 7.0. For CREPT-CTD, protein with a concentration of 16 mg mL^{-1} in the same buffer of p15RS-RPR domain was subjected to grow crystals. Crystals were grown at 18°C by vapor diffusion in hanging drop with the reservoir buffer containing 28% PEG3350, pH 5.2. Crystals were frozen in liquid nitrogen with cryoprotectant (well solution plus 20% (v/v) glycol) before data collection. Diffraction data were collected at the BL17U beam line of the Shanghai Synchrotron Research Facility. Diffraction data were indexed, integrated and scaled with the program HKL2000 [25] and the statistics are listed in Table 1.

1.6 Structural determination and refinement

The structure of p15RS-RPR domain was determined by the molecular replacement method with program PHASER [26], and the search model is Rtt103-RPR NMR structures (PDB

Table 1 Crystallographic data collection and refinement statistics

	p15RS-RPR	CREPT-CTD
Data collection		
Beamline	SSRF BL17U	SSRF BL17U
Wavelength	1.000 Å	1.000 Å
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	43.7, 47.3, 68.0	33.3, 55.7, 84.5
α, β, γ (°)	90.0, 100.2, 90.0	90.0, 92.9, 90.0
Resolution (Å)	50–2.00 (2.05–2.00)	50–2.80 (2.95–2.80)
R_{merge} (%)	8.6 (33.3)	8.4 (59.7)
$I/\sigma I$	20.1 (5.1)	6.8 (1.8)
Completeness (%)	99.8 (100)	98.6 (99.2)
Redundancy	3.7 (3.7)	2.9 (3.0)
Refinement		
Resolution (Å)	30.3–2.00	42.2–2.80
Number of reflections	18187	14346
R_{work}/R_{free} (%)	17.4/21.4	24.0/29.0
Number of atoms		
Protein	2206	2104
Water	221	
B-factors (Å^2)		
Protein	21.6	67.8
Water	30.2	
r.m.s. deviations		
Bond lengths (Å)	0.007	0.011
Bond angles (°)	0.908	1.565
Ramachandran plot		
Most favored	96.4	90.5
Additional allowed	3.6	6.7
Generously allowed	0.0	2.0
Disallowed	0.0	0.8

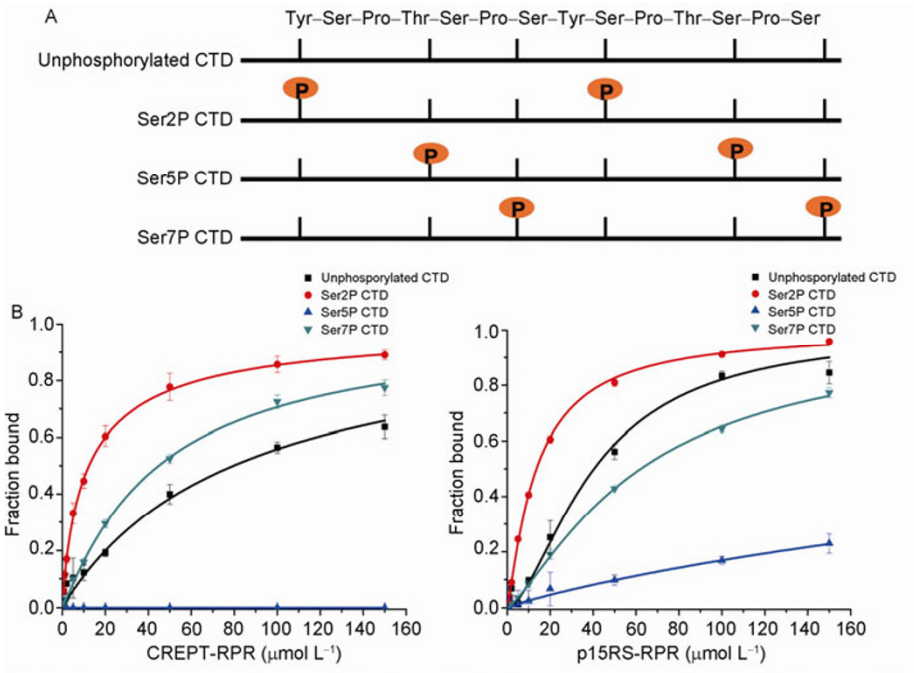
code 2KM4). After we collected the native diffraction data of CREPT-CTD, the structure of RPRD1B C-terminal domain was released (PDB code 4FLA), which is the same protein as CREPT-CTD. We therefore utilized it directly to determine the structure of CREPT-CTD using the molecular replacement method with program PHASER [26]. Structure refinement and model building were conducted with programs PHENIX and COOT, respectively [27,28]. The quality of the model was checked with program PROCHECK [29].

2 Results

2.1 CREPT and p15RS recognize RNA Pol II CTD phosphoisoforms similarly

We initially hypothesized that the opposite effects of CREPT and p15RS in the regulation of cell proliferation might result from their different binding specificity and affinity with RNA Pol II CTD. This hypothesis was supported by previous studies showing that the RPR domains from different transcription factors have different preferred CTD phosphoepitopes [20–23]. It has been documented that major phosphorylation sites on RNA Pol II CTD are on Ser2, Ser5 and Ser7 [15,23], and the smallest functional unit of RNA Pol II CTD in *in vitro* biochemical studies has been defined as two consecutive heptad peptides [21,24]. Therefore, we compared the binding affinities of CREPT-RPR and p15RS-RPR for a panel of synthetic diheptad CTD peptides with different phosphoepitopes (Figure 1A) by fluorescence polarization. In this experiment, the CREPT-RPR (residues 1–132) and p15RS-RPR (residues 1–131) with a C-terminal six histidine tag were both expressed in BL21(DE3) bacteria and purified with Nickel resin followed by size-exclusion chromatography. Both RPR domains are monomeric in solution (Figure S1 in Supporting Information).

Our fluorescence polarization analysis showed that CREPT-RPR and p15RS-RPR bound to the Ser2P RNA Pol II CTD with similar affinities of 12.3 and 13.7 μmol L^{-1}, respectively (Figure 1B). Both RRP domains bound to unphosphorylated and Ser7P RNA Pol II CTD much weaker with ~3–6 fold decreased affinities (Figure 1B). CREPT-RPR and p15RS-RPR showed the weakest binding with Ser5P RNA Pol II CTD. The p15RS-RPR bound to Ser5P RNA Pol II CTD with an affinity of less than 1 mmol L^{-1},



CTD peptide	CREPT-RPR (μmol L^{-1})	p15RS-RPR (μmol L^{-1})
Unphosphorylated CTD	79.2±59.3	40.5±8.8
Ser2P CTD	12.3±0.9	13.7±0.8
Ser5P CTD	ND	>1000
Ser7P CTD	44.9±6.3	62.2±3.3

Figure 1 Fluorescence polarization-based measurement of the binding affinities of CREPT-RPR and p15RS-RPR to diheptad CTD phosphopeptides. A, Schematic diagram of diheptad CTD peptides used in this study. Sequence is shown at the top, and serine phosphorylation sites are marked with orange circles. B, Titration curves of CREPT-RPR (top left) and p15RS-RPR (top right) domains into 0.1 μmol L^{-1} FAM-labeled unphosphorylated (black), Ser2P (red), Ser5P (blue) and Ser7P (cyan) CTD diheptad repeat peptide and calculated affinities (bottom). ND, not detectable.

and CREPT-RPR did not bind to Ser5P RNA Pol II CTD in fluorescence polarization experiment (Figure 1B). These results suggest that CREPT-RPR and p15RS-RPR exhibit similar binding specificity and affinity with RNA Pol II CTD, and Ser2P RNA Pol II CTD is the most preferred binding phosphoisoform.

2.2 The structure of p15RS-RPR

The crystal structure of p15RS-RPR was determined by the molecular replacement method and refined to a resolution of 2.0 Å with R_{work} of 17.4% and R_{free} of 21.4% (Table 1). There are two molecules in an asymmetric unit in a sub-vertical orientation (Figure S2 in Supporting Information). For each molecule, the structure identifies a compact eight-helix bundle in a right-handed super-helical arrangement (Figure 2A). The structure of p15RS-RPR is very similar to that of the yeast Rtt103-RPR domain with 1.3 Å Cα r.m.s. deviation over the entire length of the polypeptide (Figure 2B). Amino acids in the hydrophobic core of p15RS-RPR and Rtt103-RPR are highly conserved in CREPT-RPR (Figure 2C), suggesting that CREPT-RPR

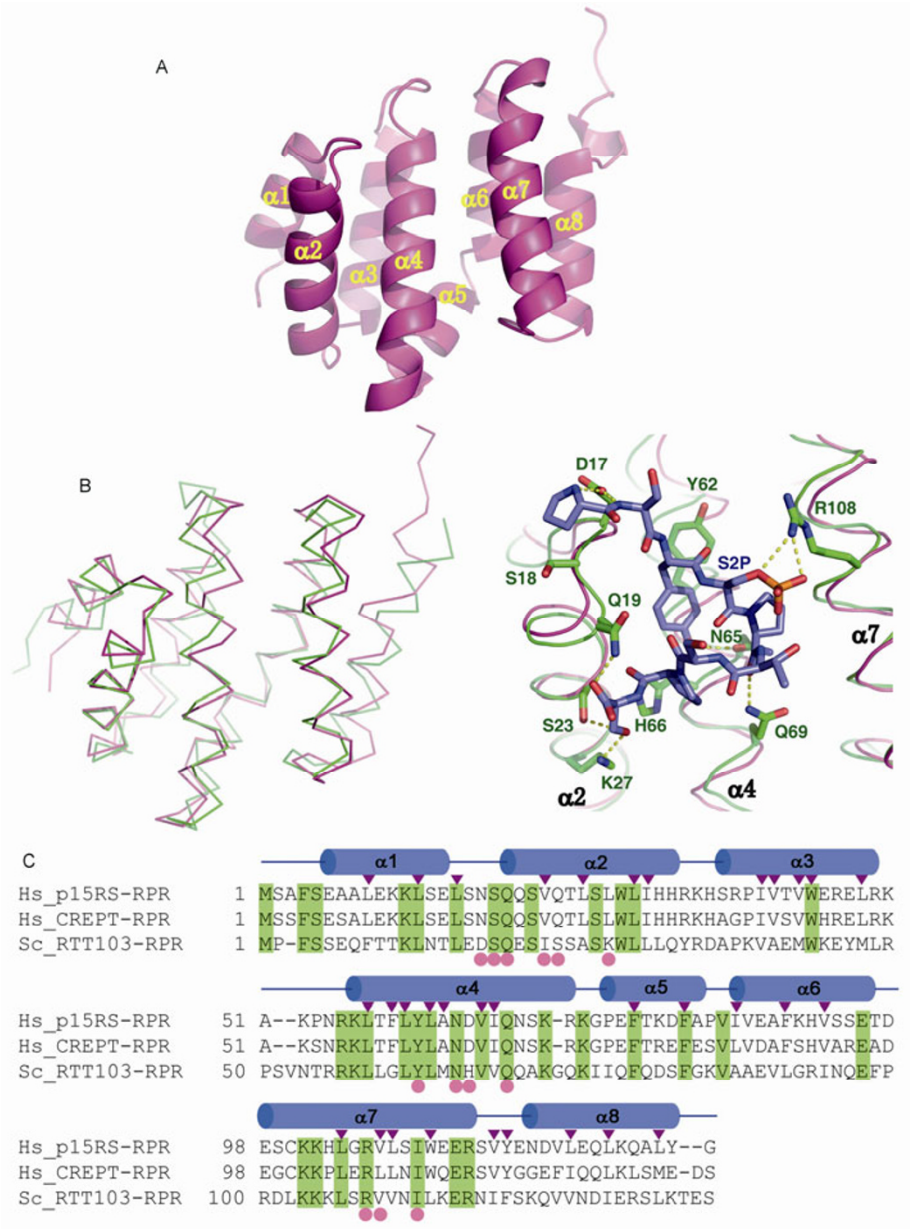


Figure 2 Structure of the N-terminal RPR domain of p15RS (p15RS-RPR). A, Ribbon diagram representation of p15RS-RPR. B, Left panel shows the structural superimposition of p15RS-RPR (purple) and Rtt103-RPR (green). Right panel shows the interactions between Rtt103-RPR (green) and Ser2P CTD diheptad peptide (blue). Hydrogen-bonding and salt-bridge interactions are represented with yellow dotted lines. C, Amino acid sequence alignment of the RPR domains of p15RS, CREPT and Rtt103. Helices are indicated as cylinders above the alignment. Residues involved in hydrophobic core packing are marked with a triangle. Residues involved in RNA Pol II CTD binding revealed in the complex structure of Rtt103-RPR with peptide are marked with a dot.

domain also adopts a similar eight-helix structure with super-helical arrangement.

A previously reported complex structure of Rtt103-RPR with a diheptad Pol II CTD peptide with two Ser2 phosphorylation positions showed that the peptide binds along helices 2, 4, and 7 of Rtt103-RPR [23], which are also the most structurally conserved helices between Rtt103-RPR and p15RS-RPR (Figure 2B). Rtt103 residues binding to the peptide are highly conserved in p15RS-RPR and CREPT-RPR (Figure 2B and C). The direct salt-bridge interaction between Arg108 in helix 7 of Rtt103 and the phosphate group of Ser2 position (Figure 2B) was suggested to be critical in determining the preference for Ser2P Pol II CTD over other phosphoisoforms [23]. This arginine residue is conserved in this position in p15RS-RPR and CREPT-RPR (Figure 2C). We expect that this arginine residue in p15RS-RPR and CREPT-RPR will also bind to the phosphate group of Ser2, which is critical in determining their binding preference among different phosphopeptides. Indeed, our fluorescence polarization analysis showed that p15RS-RPR and CREPT-RPR bind preferentially to Ser2P Pol II CTD (Figure 1).

2.3 The structure of CREPT C-terminal domain

Compared with the N-terminal RPR domain, no structure information is available for the C-terminal domain of CREPT and p15RS. We purified the CREPT-CTD (residues 177–326) expressed in BL21 (DE3) bacteria and determined its structure at 2.8 Å resolution with R_{work} of 24.0% and R_{free} of 29.0% (Table 1). CREPT-CTD is an elongated, rod-shaped dimer, with dimensions of ~30×30×140 Å (Figure 3A). Each monomer in the structure consists of residues Thr177 to Leu305, and other residues at the C-terminal tail were not built due to weak electron densities. The N-terminal region of CREPT-CTD monomer consists of three short helices α1 (Thr177-Ala190), α2 (Gly193-Ala202), and α3 (Val211-Lys216) that run approximately half the length of the rod (Figure 3A and B). After a short loop of three residues, a very long helix α4 (Lys220-Ile297) runs the entire length of the rod (Figure 3A and B). The helix α4 is continuous and has the shape of a long arch (Figure 3A). Two CREPT-CTD monomers interlock to form a symmetric dimer. The N termini of the two monomers (helix α1) are close together at the middle of the rod, and the C termini are at the opposite ends of the rod, far away from each other (Figure 3A).

2.4 Inter-chain interactions in CREPT-CTD

We further analyzed the inter-chain interactions of CREPT-CTD. Upon interlocked into a dimer, the long and symmetric binding interface between two CREPT-CRD monomers (A and B) have a very large buried surface of ~6400 Å^2 (3200 Å^2 each monomer). The two antiparallel longest α4 helices running the entire length of the dimer form a coiled-coil structure, which forms the frame of the dimer (Figure 4A). Due to the symmetric nature of the binding interface, here we only describe the interacting residues in the half of the interface. Hydrophobic residues involved in the interactions between the α4 helices include Val230(A), Leu237(A), Tyr240(A), Leu244(A), Leu248(A), Leu258(A), Tyr261(B), Leu276(B), Tyr279(B), Leu283(B), Val286(B), and Val289(B) (Figure 4A). Another significant characteristic of interactions between these two α4 helices is the surface charge complementary, including Asp231(A)-Arg290(B), Glu247(A)-Lys272(B), and Arg252(A)- Glu273 (B) salt-bridge interactions (Figure 4A). At the two far ends of the dimer, the C-terminal tail of the α4 helix further interacts with the short α3 helix of another monomer, forming

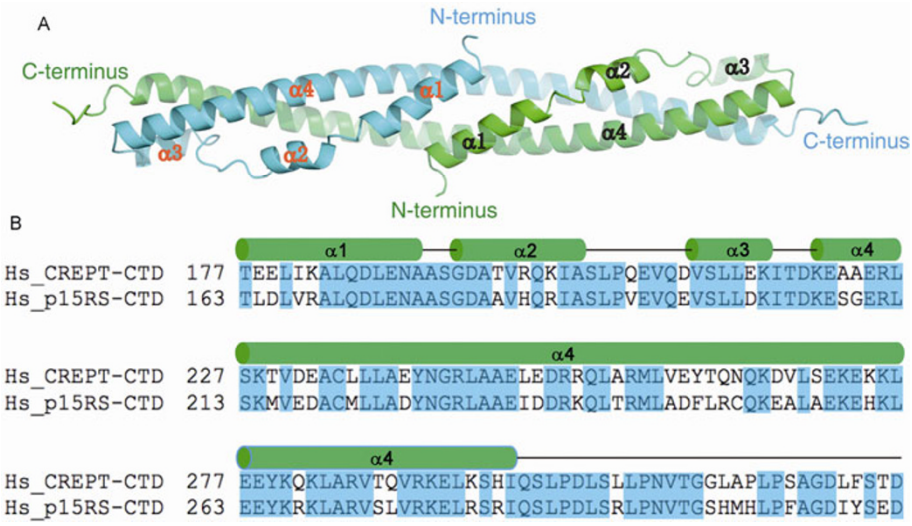


Figure 3 Structure of CREPT-CTD and sequence alignment between CREPT-CTD and p15RS-CTD. A, Ribbon diagram representation of CREPT C-terminal domain (CREPT-CTD) dimer. B, Sequence alignment of CREPT-CTD and p15RS-CTD. Conserved residues are marked with light blue background, and helices in CREPT-CTD are indicated as cylinders above the alignment.

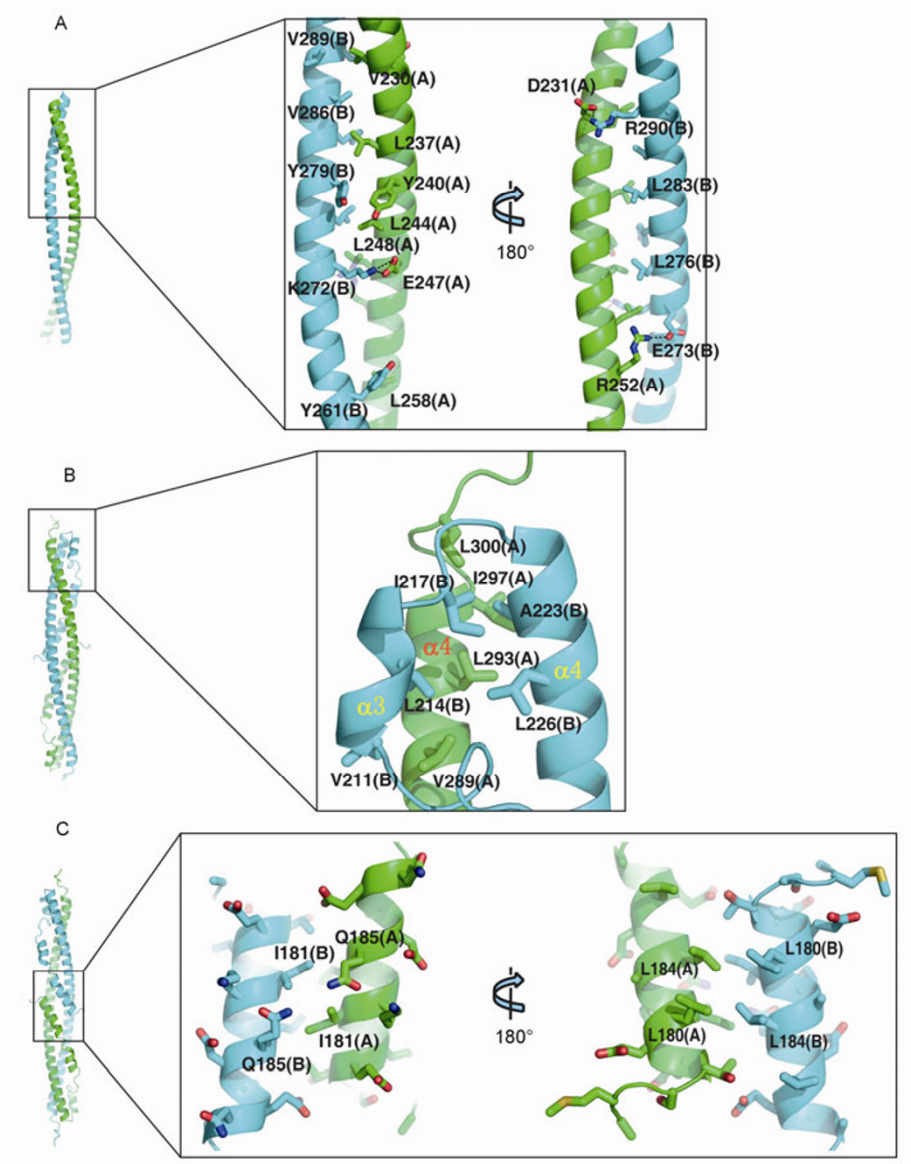


Figure 4 Inter-chain interactions in CREPT-CTD dimer. A, Packing between two α4 helices provides the frame work of the dimer. Other α helices are removed for a clear view of the interactions. Salt-bridge interactions are represented with black dotted lines. B, The interactions at the far ends of the rod-like dimer. C, Interactions between two α1 helices at the middle of the rod-like dimer.

a short triple-helix structure (Figure 4B). At this region, residues Val211(B), Leu214(B), Ile217(B), Ala223(B), and Leu226(B) form a hydrophobic cluster with Val289(A), Leu293(A), Ile297(A), and Leu300(A) (Figure 4B). At the middle of the dimer, the N-terminal α1 helices of monomers A and B have an antiparallel interaction, and the interacting hydrophobic residues include Leu180, Ile181, and Leu184 from each monomer (Figure 4C).

We proposed that the full-length CREPT might form a dimer mediated by the association of its CTD domain because the N-terminal RPR domain is monomeric in solution (Figure S1A in Supporting Information). To test this hypothesis, we measured the interaction of Flag-CREPT with Myc-CREPT expressed in 293T cells with immunoprecipitation method. In particular, we checked the effects of C-terminal truncations of Myc-CREPT on the association with full-length Flag-CREPT as the crystal structure shows that the C-terminal end of the α4 helix is involved in the CREPT-CTD dimer formation (Figure 4B). The IP results indicated that the C-terminal truncated Myc-CREPT mutants CREPT_1-285, CREPT_1-290, CREPT_1-295 and CREPT_1-300 deleting residues from 286, 291, 296 and 301 amino acid positions retained the interaction with Flag-CREPT, as strong as the interaction of full length Myc-CREPT with Flag-CREPT (Figure 5A). Deleting the C-terminal region from Val286 did not interfere with the interaction between CREPT monomers, but deleting five more residues from Gln281 to Arg285 abolished the inter-

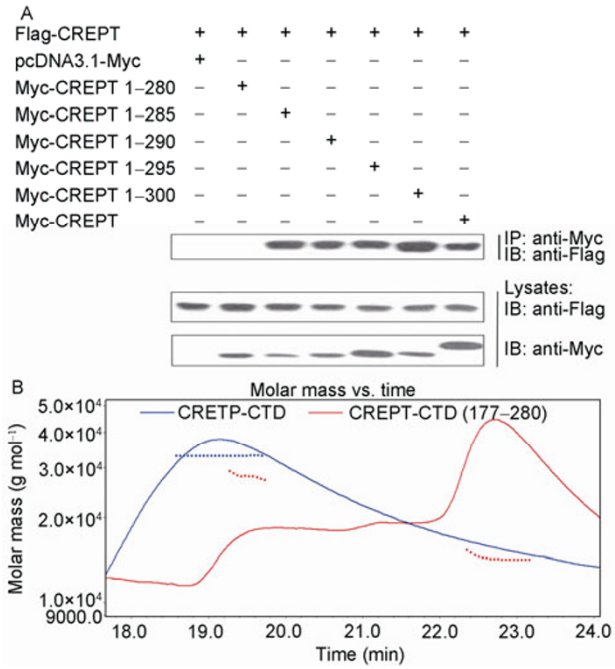


Figure 5 C-terminal deletion from Gln281 interrupts the aggregation of CREPT. A, Co-immunoprecipitation experiment between full-length CREPT-Flag and different truncation forms of CREPT-Myc. HEK293T cells were transfected with Flag-tagged full-length CREPT and different truncations of Myc-tagged CREPT. Cell lysates were immunoprecipitated using an anti-Myc antibody and the precipitants were detected with an anti-Flag antibody. B, MALS assay shows that the molecular weight of CREPT-CTD is ~33 kD, which is consistent with the dimer formation; while CREPT-CTD deleting from Gln281 (CREPT-CTD_177-280) exists mostly in monomer form (~14 kD), with a little part of dimer (~27 kD).

action (Figure 5A). Considering the interaction between full length CREPT is mediated by the C-terminal domain, we further generated a CREPT-CTD C-terminal truncation deleting residues from Gln281 (CREPT-CTD_177-280). Multiple-angle light scattering (MALS) analysis showed that most of CREPT-CTD_177-280 is monomer in solution, while the wild type CREPT-CTD is dimer (Figure 5B). This further supported that the C-terminal end of α4 helix is important for the dimer formation of CREPT-CTD, which mediates the interaction between full-length CREPT.

3 Discussion

CREPT is a recently identified oncogene that is highly expressed in different tumors [4]. It has been observed that CREPT promotes cell proliferation by enhancing the expression of cell-cycle-related genes [4]. On the other hand, the expression of p15RS, a homologous protein of CREPT, was reported to inhibit cell proliferation [3], functioning as an intrinsic inhibitor of Wnt signaling mediated transcription [5]. CREPT and p15RS both contain an N-terminal RPR domain, which recognizes RNA Pol II CTD, and a C-terminal domain with unknown function. Interestingly, both domains of CREPT and p15RS share high sequence identities. In this study, we performed biochemical and structural studies on the RPR and C-terminal domains, trying to figure out their respective roles in recognizing RNA Pol II CTD by CRETP of p15RS and explain the opposite function of CREPT and p15RS in regulating cell proliferation.

Our results demonstrated that the N-terminal RPR domain of CREPT and p15RS are very similar in binding Ser2P Pol II CTD, with higher affinity than other Pol II CTD phosphoisoforms (Figure 1). Structural determination of p15RS RPR domain and comparison with yeast Rtt103 RPR domain showed that critical residues for the interaction with Ser2P diheptapeptide are conserved in them, as well as in CREPT (Figure 2B and C), revealing a structural basis for the binding preference to Ser2P diheptapeptide by CREPT and p15RS. We previously have shown that CREPT is able to bind RNA Pol II at the promoter and terminator regions of genes and proposed that it recycles the Pol II to the 5′ promoter region by a chromatin loop [4]. RNA Pol II at the late stage of transcription is modified by phosphorylation at Ser2 positions of RNA Pol II CTD, while Ser5P is the most dominating modification state of CTD before the late stage of transcription [10,13,15,17]. Our fluorescence polarization analysis showed that CREPT-RPR did not bind to the Ser5P diheptapeptide, even weaker than the binding with the unphosphorylated diheptapetpide (Figure 1). Therefore, we proposed that CREPT might selectively associate with Ser2P CTD of RNA Pol II at the 3′ terminal end of a transcribing gene. Whether the preferred interaction of CREPT with Ser2P CTD of RNA Pol II would facilitate the formation of a chromatin loop by the association with 5′ promoter region remained to be examined. Currently, we are not clear about the molecular basis of the association of CREPT with 5′ promoter region of a gene. Since CREPT failed to associate with Ser5P CTD of RNA Pol II, which remains at a high level in the promoter region during the transcription initiation, we speculate that other proteins that could bind to DNA might associate with Ser5P CTD of RNA Pol II and CREPT simultaneously to regulate the status of RNA Pol II in the promoter region.

The RPR domain has also been found in several RNA processing and transcription termination factors other than CREPT and p15RS [20,22,23]. The RPR containing proteins additionally have other domains responsible for protein-protein and protein-RNA interactions such as Pcf11, Rtt103, and SCAF8. However, CREPT and p15RS contain only the RPR domain, except their conserved C-terminal domain without a high sequence homology with other proteins. The structure of CREPT-CTD shows that C-terminal domain of CREPT forms a very long rod-like dimer, each monomer adopting a coiled-coil structure (Figure 3A). We analyzed the interactions between two CREPT-CTD domains and found that the C-terminal residues from Gln281 to Arg285 are critical for the dimer formation of the

CREPT-CTD in solution (Figure 5B), as well as for the interaction between full length CREPT monomers (Figure 5A) when they associate together to form a dimer or other aggregated forms in the cell.

We searched the Dali server and identified some proteins containing a similar coiled-coil structure as the CREPT-CTD but with diverse functions (data not shown). The coiled-coil structure is a popular structural motif found in many proteins, structural similarity therefore is not able to provide further insights into the function of CTD in CREPT and p15RS. We propose that CREPT may form a dimer in the association with RNA Pol II. It is speculated that the CREPT-CTD dimer might directly enhance the interaction of CREPT with RNA Pol II. The N-terminal RPR domain functions as a monomer, supported by both biochemical data and the crystal structure (Figure S1A in Supporting Information; Figure 2A). The complex structure of Rtt103-RPR domain with Ser2P diheptapeptide suggests that the RPR domain monomer has one binding site for the peptide, especially for the Ser2P [23]. The RNA Pol II CTD consists of up to 52 heptapeptide repeats [11]. The full-length CREPT holds a very stable dimer in its C-terminal domain, and two N-terminal RPR domains are linked to the C-terminal dimer via a long linker (Figure 6). The flexibility of the linker enables the two RPR domains to interact with the RNA Pol II CTD with Ser2P individually with a similar affinity. The dimeric CREPT complex that has two Pol II CTD binding sites is expected to increase the binding avidity with Pol II CTD.

The functional role of the RPR domain in direct association with RNA Pol II CTD has been studied [20–23]. Here we firstly report the dimer structure of CREPT C-terminal domain and propose that one function of CREPT-CTD dimer is to increase binding strength of CREPT with Pol II CTD (Figure 6). We expect that the RPR and C-terminal domains in p15RS would have similar functions because the highly conserved primary amino acid sequences in these two domains (Figure S3 in Supporting Information) and our MALS data showed that p15RS-CTD is dimer in solution (Figure S4 in Supporting Information). This raises an interesting question still not answered: what is the molecular mechanism underlying the opposite function in regulating cell proliferation by CREPT and p15RS? Besides the proposed role of increasing binding avidity, we cannot exclude that the similar C-terminal domains of CREPT and p15RS bind to different protein partners that would determine their different functions *in vivo*. Another possibility resides in the linker region connecting the RPR to the C-terminal domain (Figure S3 in Supporting Information), which demonstrates quite diverse sequences in CREPT and p15RS.

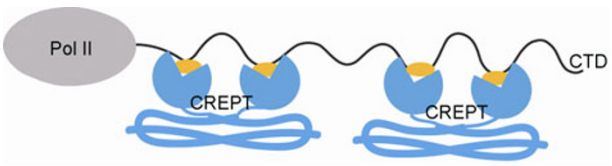


Figure 6 A model of interaction between CREPT or p15RS dimer and RNA Pol II CTD.

PDB Deposition

The coordinates and diffraction data have been deposited into the Protein Data Bank with accession codes of 4NAC for p15RS-RPR and 4NAD for CREPT-CTD.

We thank Dr. Wang JiaWei for assistance with structure determination, and Dr. He JianHua and other staff members at Shanghai Synchrotron Research Facility (SSRF) beam line BL17U for help with data collection. This work was supported by Ministry of Science and Technology (2010CB912402 and 2011CB910502), Ministry of Health (2012ZX1000-1009) and the Fok Ying Tung Education Foundation to Wang XinQuan and the National Natural Science Foundation of China (81230044, 81372167 and 31071225) and the Tsinghua Science Foundation (20121080018) to Chang ZhiJie.

1 Hanahan D, Weinberg RA. The hallmarks of cancer. Cell, 2000, 100: 57–70
2 Gordon GJ, Rockwell GN, Jensen RV, Rheinwald JG, Glickman JN, Aronson JP, Pottorf BJ, Nitz MD, Richards WG, Sugarbaker DJ, Bueno R. Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. Am J Pathol, 2005, 166: 1827–1840
3 Liu J, Liu H, Zhang X, Gao P, Wang J, Hu Z. Identification and characterization of P15RS, a novel P15(INK4b) related gene on G1/S progression. Biochem Biophys Res Commun, 2002, 299: 880–885
4 Lu D, Wu Y, Wang Y, Ren F, Wang D, Su F, Zhang Y, Yang X, Jin G, Hao X, He D, Zhai Y, Irwin DM, Hu J, Sung JJ, Yu J, Jia B, Chang Z. CREPT accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. Cancer Cell, 2012, 21: 92–104
5 Wu Y, Zhang Y, Zhang H, Yang X, Wang Y, Ren F, Liu H, Zhai Y, Jia B, Yu J, Chang Z. p15RS attenuates Wnt/{beta}-catenin signaling by disrupting {beta}-catenin.TCF4 Interaction. J Biol Chem, 2010, 285: 34621–34631
6 Zhang X, Cao Q, Liu X, Liu S, Wang J, Sun S, Wang O, Tian Z, Liu H, Kuang J, Zhang W. Cellular and molecular evidence for malignancy-inhibitory functions of p15RS. Cell Cycle, 2012, 11: 1988–1998
7 Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature, 2004, 432: 517–522
8 Srividya I, Tirupataiah S, Mishra K. Yeast transcription termination factor Rtt103 functions in DNA damage response. PLoS ONE, 2012, 7: e31288
9 Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. Cell, 2002, 108: 501–512
10 Hirose Y, Manley JL. RNA polymerase II and the integration of nuclear events. Genes Dev, 2000, 14: 1415–1429
11 Corden JL. Tails of RNA polymerase II. Trends Biochem Sci, 1990, 15: 383–387
12 Dahmus ME. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J Biol Chem, 1996, 271: 19009–19012
13 Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev, 2000, 14: 2452–2460
14 Ahn SH, Kim M, Buratowski S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol Cell, 2004, 13: 67–76
15 Egloff S, Murphy S. Cracking the RNA polymerase II CTD code. Trends Genet, 2008, 24: 280–288

16 Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev, 2006, 20: 2922–2936
17 Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell, 1999, 3: 405–411
18 McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev, 1997, 11: 3306–3318
19 Cho EJ, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev, 1997, 11: 3319–3326
20 Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature, 2004, 430: 223–226
21 Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol, 2005, 12: 144–151
22 Becker R, Loll B, Meinhart A. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. J Biol Chem, 2008, 283: 22659–22669
23 Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol, 2010, 17: 1195–1201
24 Stiller JW, Cook MS. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryot Cell, 2004, 3: 735–740
25 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter C W Jr., ed. Methods in Enzymology. Academic Press, 1997. 307–326
26 McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr, 2007, 63: 32–41
27 Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr, 2002, 58: 1948–1954
28 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr, 2004, 60: 2126–2132
29 Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst, 1993, 26: 283–291

Supporting Information

Figure S1 Gel-filtration of CREPT-RPR (A) and p15RS-RPR (B). The coordinate graph on the top right of each figure is a plot of log molecular weight against elution volume of the column. Gel filtration standards (Bio-Rad) used in the experiment were vitamin B12 (1.3 kD), horse myoglobin (17 kD), chicken ovalbumin (44 kD), bovine gamma-globulin (158 kD), and bovine thyroglobulin (670 kD). Theoretical molecular weight of CREPT-RPR and p15RS-RPR monomer is 15.283 and 15.22 kD, respectively. Calculated molecular weight from gel filtration of CREPT-RPR and p15RS-RPR is 14.092 and 14.628 kD, indicating that CREPT-RPR and p15RS-RPR are monomers.

Figure S2 Two p15RS-RPR molecules in an asymmetric unit are in a sub-vertical orientation.

Figure S3 Sequence alignment of human p15RS and CREPT. Red box indicates the long loop linking N-terminal RPR domain and C-terminal domain (CTD) which shows high diversity. Residues 281 to 285 important for CREPT dimer formation are indicated.

Figure S4 MALS analysis shows that the molecular weight of p15RS-CTD in solution is around 35 kD (red curve), indicating that it is a dimer in solution. The molecular weight of CREPT-CTD dimer (purple curve) in solution is also around 35 kD.

The supporting information is available online at life.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.